SANDIA REPORT

SAND94—2642 • UC—905
Unlimited Release
Printed November 1994

C.1

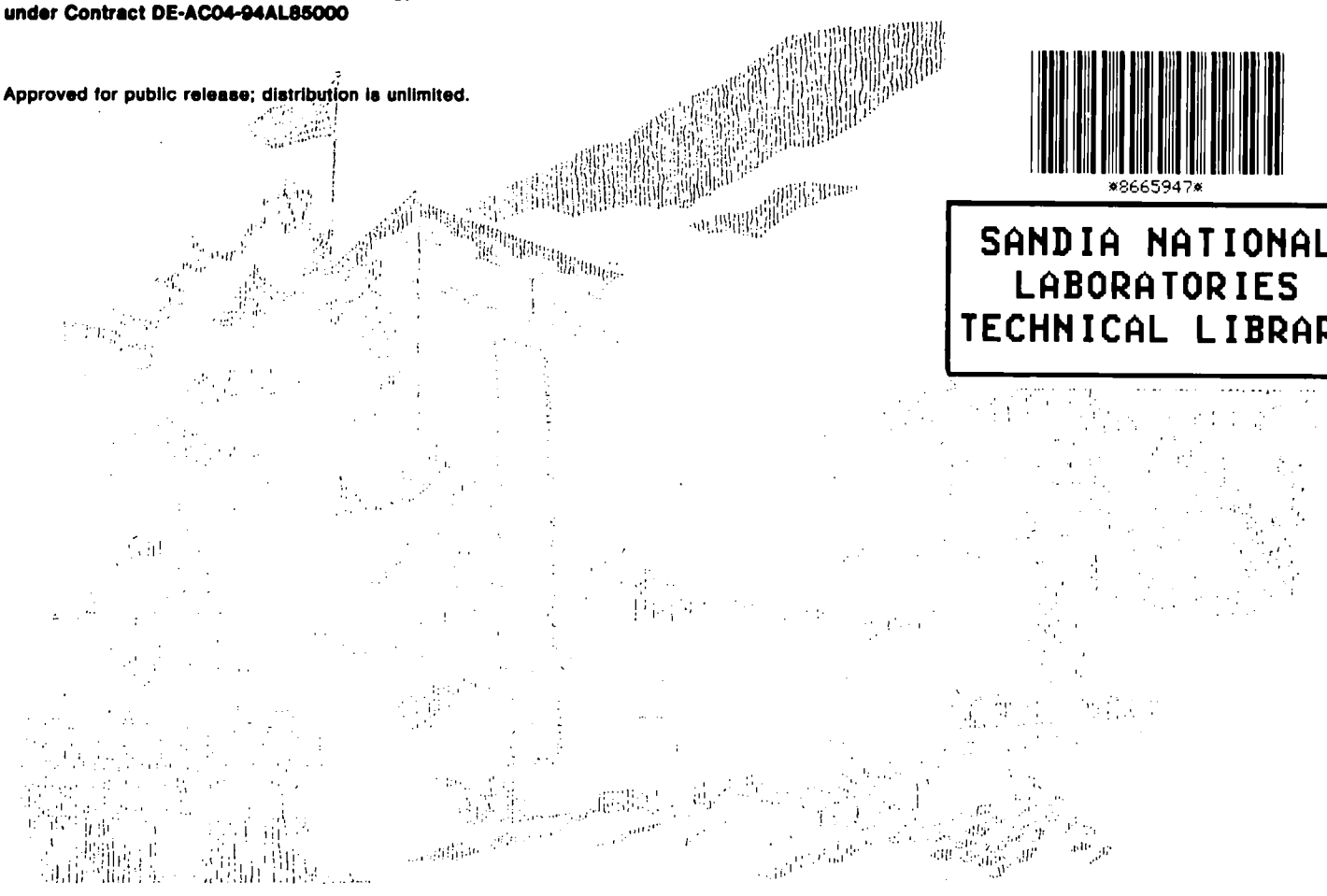
MICROFICHE

Simultaneous Inference in Multivariate Calibration

Edward V. Thomas

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
for the United States Department of Energy
under Contract DE-AC04-94AL85000

Approved for public release; distribution is unlimited.



28p.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831

Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
US Department of Commerce
5285 Port Royal RD
Springfield, VA 22161

NTIS price codes
Printed copy: A03
Microfiche copy: A06

SAND 94-2642 Distribution
Unlimited Release Category UC-905

Simultaneous Inference in Multivariate Calibration

Edward V. Thomas
Statistics and Human Factors Department
Sandia National Laboratories

ABSTRACT

A set of q responses, $\mathbf{y}=(y_1, y_2, ..., y_q)^T$, is related to a set of p explanatory variables, $\mathbf{x}=(x_1, x_2, ..., x_p)^T$, through the classical linear model, $\mathbf{y}^T=\mathbf{a}+\mathbf{x}^T\mathbf{B}+\mathbf{e}^T$. The parameters, $\mathbf{a}$ and $\mathbf{B}$, are estimated during calibration using a *training set*. The fitted calibration model that follows is then used repeatedly on a number of new observations where $\mathbf{y}$ is observed and $\mathbf{x}$ is to be inferred. This procedure is often referred to as prediction (or inverse prediction).

The prediction procedure can be viewed as parameter estimation in *errors-in-variables* regression (see Thomas 1991). By using the errors-in-variables connection and assuming normally distributed measurement errors, the *maximum likelihood estimates* of the new $\mathbf{x}$'s can be obtained either individually or jointly. The limiting ($q\rightarrow\infty$) normal distribution of the maximum likelihood estimates of the new $\mathbf{x}$'s, obtained jointly, can be used to construct approximate simultaneous confidence regions for the new $\mathbf{x}$'s. In situations where the new observations are numerous and well dispersed from the center of the training set, the uncertainty in $\mathbf{a}$ and $\mathbf{B}$ is substantial, and the specificity of the responses is poor, joint estimation can improve significantly on individual estimation, which is the traditional approach.

1. INTRODUCTION

Recently, calibration involving a multidimensional response has received significant attention (e.g., see Brown 1982, and Næs 1985). Osborne's (1991) recent review of calibration includes discussion of the wide variety of approaches to calibration involving a multidimensional response. Multivariate calibration is used in a number of applications, often in analytical chemistry. Many of these applications involve a blending of computer technology and specialized instrumentation (e.g., spectrometers) that result in response variables of a very high dimension. The discussion presented here is particularly relevant for this situation. Martens and Næs (1989) and references therein provide a number of examples with a high-dimensional response variable.

Because of the cost of developing a calibration model, a single fitted calibration model is often used repeatedly on a number of new observations (multiple-use calibration) to predict some characteristic of interest. For example, in analytical chemistry a calibration model is often constructed to predict some characteristic (e.g., chemical concentration) in a batch or stream comprising a number of new specimens (observations). Within the batch (or block within a stream), specimens are often prepared together. After the specimens are prepared, they are subjected to measurement by the appropriate analytical instrument (e.g., spectrometer). Prediction follows the acquisition of the instrumental measurements.

The purpose of this article is to demonstrate the potential advantage of using the responses of the new observations jointly for prediction. This differs from the traditional approach where prediction of the characteristic of interest for an individual new observation involves only the multivariate response associated with that single new observation. The development of the joint prediction approach follows from Thomas (1991) where prediction was developed as an estimation problem in errors-in-variables regression. Maximum likelihood estimation (assuming normally distributed errors in the responses) was proposed as an alternative to least-squares estimation because of the consistency of the maximum likelihood estimator (MLE) with respect to the number of relevant response variables. In this article, the underlying errors-in-variables model is used to modify the MLE to include information inherent in the responses from all new observations.

Calibration consists of two distinct steps; calibration and prediction. In the case of multivariate linear calibration (considered exclusively here), the calibration step consists of estimating $\mathbf{a} = (a_1, a_2, \ldots, a_q)$ and $\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_q)$ in the model

$\mathbf{Y} = \mathbf{X}_{\mathbf{f}}\mathbf{B}_{\mathbf{f}} + \mathbf{E}$, where [1.1]

$\mathbf{Y}=(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_n)^T$ is the n × q matrix of responses,

$$\mathbf{X}_f=\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}^T \text{ is a } n\times(p+1) \text{ matrix of constants, and}$$

$$\mathbf{B}_f=\begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_q \end{bmatrix} \text{ is a } (p+1)\times q \text{ matrix of unknown parameters,}$$

Here, q is the number of response variables, n is the number of calibration samples, and p is the number of explanatory variables. Note that this model represents *controlled calibration*, as the $\mathbf{x}_i$ are assumed to be fixed (see Brown 1982). For notational simplicity it is assumed that the $\mathbf{x}_i$ are centered, so that $\sum_{i=1}^{n}\mathbf{x}_i=\mathbf{0}$.

Further, it will be assumed that the rows of the matrix of random errors, $\mathbf{E}=(\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n)^T$, are multivariate normal with 1. $\mathbf{E}(\mathbf{e}_i)=\mathbf{0}$, 2. $\mathbf{E}(\mathbf{e}_i\mathbf{e}_j^T)=\mathbf{0}$ for $i\neq j$, and 3. $\mathbf{E}(\mathbf{e}_i\mathbf{e}_i^T)=\sigma^2\mathbf{I}_q$. More generally, we could allow $\mathbf{E}(\mathbf{e}_i\mathbf{e}_i^T)\propto\mathbf{V}$, with $\mathbf{V}$ known, followed by the appropriate transformation of the rows of $\mathbf{Y}$ to achieve condition 3. In this article, it will be assumed that this transformation, if necessary, has been performed. In general, note that if $\mathbf{V}$ has a completely arbitrary structure, it may not be possible to obtain a proper estimate of $\mathbf{E}(\mathbf{e}_i\mathbf{e}_i^T)=\sigma^2\mathbf{I}_q$ since we often have n much smaller than q. However, in chemical applications involving optical methods, the ordered nature of the responses (embodied in a spectrum) can often give rise to a structured $\mathbf{V}$ involving only a few parameters (e.g., see Denham and Brown 1991 and Thomas 1991). Although in practice $\mathbf{V}$ is unlikely to be known, transformation of $\mathbf{Y}$ based on an estimate of $\mathbf{V}$ will approximately satisfy condition 3. This article does not address the consequences of uncertainty in the estimate of $\mathbf{V}$.

By using least-squares regression, an estimate of $\mathbf{B}_f$ is

$$\hat{\mathbf{B}}_f=\begin{bmatrix} \hat{a}_1 & \hat{a}_2 & \cdots & \hat{a}_q \\ \hat{\mathbf{b}}_1 & \hat{\mathbf{b}}_2 & \cdots & \hat{\mathbf{b}}_q \end{bmatrix}=(\mathbf{X}_f^T\mathbf{X}_f)^{-1}\mathbf{X}_f^T\mathbf{Y}. \qquad [1.2]$$

The errors in estimating $\mathbf{B}_f$ are denoted by

$$\Delta_f=\hat{\mathbf{B}}_f-\mathbf{B}_f=\begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_q \\ \delta_1 & \delta_2 & \cdots & \delta_q \end{bmatrix}.$$

It follows that $(\gamma_t,\delta_t^T)^T \overset{iid}{\sim} \mathbf{Normal}(\mathbf{0},\sigma^2(\mathbf{X}_f^T\mathbf{X}_f)^{-1})$.

In the prediction phase, new samples are obtained where, for each sample, only the q-vector, $\mathbf{y}$, is observed. The objective is to estimate the p-vector, $\mathbf{x}$, that underlies each of these new samples. Here, no distribution of the new $\mathbf{x}$'s will be assumed. The parameter estimates, $\hat{\mathbf{B}}_{\dagger}$, obtained earlier in the calibration phase, are repeatedly used in the prediction phase. There are two sources of error when estimating the $\mathbf{x}$'s for the new samples. One source is the error in the calibration, $\Delta_{\dagger}$. The other source is the measurement errors associated with the new $\mathbf{y}$'s. The common calibration error ($\Delta_{\dagger}$) introduces correlation among the errors in estimating the various new $\mathbf{x}$'s.

A number of authors (e.g., see Scheffé 1973; Carroll, Spiegelman, and Sacks 1988; and Mee, Eberhardt, and Reeve 1991) have addressed the difficult problem of simultaneous inference in univariate calibration ($q = p = 1$). This article develops simultaneous inference in multivariate calibration by exploiting the errors-in-variables representation of the prediction phase outlined by Thomas (1991). When there are more response variables than explanatory variables (i.e. $q > p$), the overdetermined nature of the prediction phase makes inference in calibration with many variables a somewhat easier problem than the analogous problem in univariate calibration where prediction is completely determined.

The remainder of this paper consists of the following. Section 2 formulates the prediction problem as an estimation problem in errors-in-variables regression. Sections 3 and 4 describe individual and joint maximum likelihood estimation of the new $\mathbf{x}$'s. The limiting ($q \to \infty$) distributions of the *individual* and *joint* MLE's are given. An asymptotic comparison of the two estimation methods is presented in Section 5. In Section 6, the two estimation procedures are compared using simulations. Section 7 contains a short conclusion.

2. ERRORS-IN-VARIABLES MODEL FOR ESTIMATING $\mathbf{x}_0$'s

During prediction, r new observations are obtained where only the q-vectors, $\mathbf{y}^{(j)}$, in $\mathbf{Y}_0 = \left(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(r)}\right) = \left(\mathbf{y}_{(1)}^T, \mathbf{y}_{(2)}^T, \ldots, \mathbf{y}_{(q)}^T\right)^T$, are observed. The p-vectors, $\mathbf{x}^{(j)}$, in $\mathbf{X}_0 = \left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(r)}\right)$, are to be estimated. The parameter estimates obtained earlier in the calibration phase, $\hat{\mathbf{B}}_{\dagger}$, are used for this purpose.

The prediction of the r new observations can be parametrized in terms of a *linear functional model* (e.g., see Fuller 1987). The model is $\mathbf{u}_t = \mathbf{b}_t^T \mathbf{X}_0 + a_t \mathbf{1}$ (or, $\mathbf{z}_t = \mathbf{u}_t - a_t \mathbf{1} = \mathbf{b}_t^T \mathbf{X}_0$), $\hat{\mathbf{b}}_t = \mathbf{b}_t + \boldsymbol{\delta}_t$, $\hat{a}_t = a_t + \gamma_t$, and $\mathbf{y}_{(t)} = \left(y_{t1}, y_{t2}, \ldots, y_{tr}\right) = \mathbf{u}_t + \mathbf{e}_{(t)}$, for $t = 1, 2, \ldots, q$. The underlying linear functional relationship is $\mathbf{z}_t = \mathbf{b}_t^T \mathbf{X}_0$. The observed values of the response

variables, in $\mathbf{Y_0}$, contain the unobservable true values of the responses $(\mathbf{u_t})$ and measurement errors, $\mathbf{e_{(t)}} = (e_{t1}, e_{t2}, \ldots, e_{tr})$, which are assumed independent across the r new observations. Let $\mathbf{Z_t} = \left(Z_{t1}, Z_{t2}, \ldots, Z_{tr}\right) = \mathbf{z_t} + \mathbf{e_{(t)}} - \gamma_t \mathbf{1}$ be the observed value of $\mathbf{z_t} = \left(z_{t1}, z_{t2}, \ldots, z_{tr}\right)$. That is $\mathbf{Z_t} = \mathbf{y_{(t)}} - \hat{a}_t \mathbf{1}$. Also, let $\epsilon_{tj} = (e_{tj} - \gamma_t, \delta_t^T)^T$. From earlier assumptions, it can be shown that for fixed j, $\epsilon_{tj} \overset{iid}{\sim} \mathbf{Normal}(\mathbf{0}, \Sigma)$ for t = 1, 2,..., q, where $\Sigma = \sigma^2 \mathbf{\Omega}$ and

$$\mathbf{\Omega} = \begin{bmatrix} \Omega_{11} & \mathbf{0} \\ \mathbf{0} & \Omega_{22} \end{bmatrix} = (\mathbf{X_1^T X_1})^{-1} + \mathrm{diag}(1, 0, 0, \ldots, 0).$$

Further, $\epsilon_{t\cdot} = (e_{t1} - \gamma_t, e_{t2} - \gamma_t, \ldots, e_{tr} - \gamma_t, \delta_t^T)^T \overset{iid}{\sim} \mathbf{Normal}(\mathbf{0}, \mathbf{\Sigma_a})$ for t = 1, 2,..., q, where $\mathbf{\Sigma_a} = \sigma^2 \mathbf{\Omega_a}$,

$$\mathbf{\Omega_a} = \begin{bmatrix} \Omega_{11} & \Omega_{\gamma\gamma} & \cdots & \Omega_{\gamma\gamma} & \mathbf{0} \\ \Omega_{\gamma\gamma} & \Omega_{11} & & & \\ \vdots & & \ddots & \Omega_{\gamma\gamma} & \vdots \\ \Omega_{\gamma\gamma} & \cdots & \Omega_{\gamma\gamma} & \Omega_{11} & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \Omega_{22} \end{bmatrix}, \text{ and } \Omega_{\gamma\gamma} = \mathrm{Var}(\gamma_t).$$

Notice that the set of estimated model parameters $\{\hat{a}_t, \hat{\mathbf{b}}_t\}$ with associated errors $\{\gamma_t, \delta_t\}$ remain constant over all r new observations. This introduces correlation among the errors associated with maximum likelihood (and other) estimates of the r p-vectors of $\mathbf{X_0}$. We will consider estimation of the columns of $\mathbf{X_0}$ by maximum likelihood both individually (one column at a time) and jointly.

3. INDIVIDUAL MAXIMUM LIKELIHOOD ESTIMATION

In many calibration applications, the columns of $\mathbf{X_0}$ are estimated individually. For example, in analytical chemistry, it is common practice to estimate the chemical characteristic of interest for a new sample without regard to the responses of the other new samples. This practice is prevalent despite the fact that instrumental measurements for new samples are often acquired in a batch.

Given the knowledge of the distribution of ϵ_{tj}, the individual MLE of $\mathbf{x^{(j)}}$ (and B) can be obtained by maximizing the log likelihood,

$$-(n/2) \cdot \log\left|2\pi\sigma^2\mathbf{\Omega}\right| - (2\sigma^2)^{-1} \sum_{t=1}^{q} \left((Z_{tj}, \hat{\mathbf{b}}_t^T) - (z_{tj}, \mathbf{b}_t^T) \right) \mathbf{\Omega}^{-1} \left((Z_{tj}, \hat{\mathbf{b}}_t^T) - (z_{tj}, \mathbf{b}_t^T) \right)^T,$$

with respect to $\mathbf{x}^{(j)}$ and $\mathbf{B}$ (e.g., see Fuller 1987). Recall that z_{tj} is defined explicitly by the underlying linear functional relationship, $z_{tj} = \mathbf{b}_t^T \mathbf{x}^{(j)}$. If Ω is nonsingular and $q \geq p$, the individual MLE of $\mathbf{x}^{(j)}$ is

$$\tilde{\mathbf{x}}^{(j)} = (\mathbf{M}_{22} - \hat{\lambda}\Omega_{22})^{-1}\mathbf{M}_{21}^{(j)}, \text{ where} \qquad [3.1]$$

$\mathbf{M}_{22} = q^{-1}\sum_{t=1}^{q} \hat{\mathbf{b}}_t \hat{\mathbf{b}}_t^T$, $\mathbf{M}_{21}^{(j)} = q^{-1}\sum_{t=1}^{q} \hat{\mathbf{b}}_t z_{tj}$, and $\hat{\lambda}$ is the smallest root of $|\mathbf{M}^{(j)} - \lambda\Omega| = 0$.

$$\mathbf{M}^{(j)} = \begin{bmatrix} M_{11}^{(j)} & \mathbf{M}_{12}^{(j)} \\ \mathbf{M}_{21}^{(j)} & \mathbf{M}_{22} \end{bmatrix} = q^{-1}\sum_{t=1}^{q} \left(z_{tj}, \hat{\mathbf{b}}_t^T\right)^T \left(z_{tj}, \hat{\mathbf{b}}_t^T\right).$$

Note that Brown and Sundberg (1987) give an approach for obtaining the MLE of $\mathbf{x}^{(j)}$ when $\mathbf{E}(\mathbf{e}_i \mathbf{e}_i^T)$ is arbitrary and unknown. However, the development of the MLE when $\mathbf{E}(\mathbf{e}_i \mathbf{e}_i^T)$ is completely arbitrary and unknown requires that the number of calibration samples (n) exceeds $q + p$. This condition is not likely to be met in situations in which q is very large (perhaps 1000 or more). If $\mathbf{E}(\mathbf{e}_i \mathbf{e}_i^T)$ is assumed known within a scale constant, then the approach given by Brown and Sundberg (1987) will result in the expression given in equation (3.1).

The purpose of $\hat{\lambda}\Omega_{22}$ is to shrink $\mathbf{M}_{22}$ towards $\mathbf{m}_{22} = q^{-1}\sum_{t=1}^{q} \mathbf{b}_t \mathbf{b}_t^T$. By assuming that $\bar{\mathbf{b}} = \lim_{q\to\infty} q^{-1}\sum_{t=1}^{q} \mathbf{b}_t$ and $\bar{\mathbf{m}}_{22} = \lim_{q\to\infty} q^{-1}\sum_{t=1}^{q} \mathbf{b}_t \mathbf{b}_t^T$ is positive definite, $\tilde{\mathbf{x}}^{(j)}$ is strongly consistent with respect to q (see Amemiya and Fuller 1984). Note that in many chemical applications p is relatively small (≤ 5), while q can be very large (sometimes >1000). Thus, the limiting distribution of $\tilde{\mathbf{x}}^{(j)}$ with respect to q (to be discussed) has particular relevance.

Coincidentally, the maximum likelihood estimates of $\left(z_{tj}, \mathbf{b}_t^T\right)^T$ modified by the j^{th} new observation is $\left(\tilde{z}_{tj}, \tilde{\mathbf{b}}_t^{(j)T}\right) =$

$$\left(Z_{tj}, \hat{\mathbf{b}}_t^{(j)T}\right) - \left(Z_{tj} - \hat{\mathbf{b}}_t^{(j)T}\tilde{\mathbf{x}}^{(j)}\right)\left[\left(1, -\tilde{\mathbf{x}}^{(j)T}\right)\Omega\left(1, -\tilde{\mathbf{x}}^{(j)T}\right)^T\right]^{-1}\left(1, -\tilde{\mathbf{x}}^{(j)T}\right)\Omega. \qquad [3.2]$$

(See e.g., Fuller 1987, page 125). Note that unique estimates of $\mathbf{b}_t^T$ are obtained for each new sample. Also, note that because of the error structure we are not able to obtain separate estimates of a_t and y_{tj} from $\tilde{z}_{tj}$.

In general, with arbitrary p, Σ, and q, the exact distribution of $\tilde{\mathbf{x}}_0^{(j)}$ is unknown. However, with the earlier assumptions on $\bar{\mathbf{b}}$ and $\bar{\mathbf{m}}_{22}$, it can be shown that as $q \to \infty$, $\Gamma_j^{-\frac{1}{2}}(\tilde{\mathbf{x}}^{(j)} - \mathbf{x}^{(j)}) \xrightarrow{\mathcal{L}} \text{Normal}(\mathbf{0}, \mathbf{I})$, where

$$\Gamma_j = \frac{\sigma^2}{q}\psi_j \mathbf{m}_{22}^{-1} + \frac{\sigma^4}{q}\mathbf{m}_{22}^{-1}(\Omega_{22}\psi_j - \Lambda_j \Lambda_j^T)\mathbf{m}_{22}^{-1}, \qquad [3.3]$$

$\psi_j=(1,-\mathbf{x}^{(j)^T})\Omega(1,-\mathbf{x}^{(j)^T})^T$, $\mathbf{A}_j=-\Omega_{22}\mathbf{x}^{(j)}$, and $\Gamma_j^{-\frac{1}{2}}$ is an upper triangular matrix, such that $(\Gamma_j^{-\frac{1}{2}})^T\Gamma_j^{-\frac{1}{2}}=\Gamma_j^{-1}$. This expression can be used as the basis for deriving a confidence region for an individual $\mathbf{x}^{(j)}$ (see Thomas 1991). Note that the assumption about $\bar{m}_{22}$ implies the need for *relevant* responses (i.e. $\mathbf{b}_t \neq 0$) and *specific* responses (i.e. the rows of $\mathbf{B}=(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_q)$ are different).

In practice, neither σ^2, $\mathbf{x}^{(j)}$, or m_{22}, which are elements of Γ_j, will be known. An estimator of Γ_j, $\hat{\Gamma}_j$, following that given by Fuller (1987), page 130, and Thomas (1991) can be obtained by replacing σ^2, $\mathbf{x}^{(j)}$, or m_{22} in the expression for Γ_j with appropriate estimates. An estimate of σ^2 is $\hat{\sigma}^2=\Big((n+1)\cdot(q-p)-n\Big)^{-1}\cdot\Big\{\Big(n\cdot(q-p-1)\Big)\hat{\sigma}_c^2+(q-p)\,\hat{\sigma}_p^2\Big\}$, where

$\hat{\sigma}_c^2=\Big(n\cdot(q-p-1)\Big)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{q}\Big(Y_{it}-\mathbf{X}_i\hat{\mathbf{b}}_t-\hat{a}_t\Big)^2$ is an estimate of $\hat{\sigma}^2$ obtained from the calibration phase (Y_{it} is the it^{th} element of $\mathbf{Y}$ and $\mathbf{X}_i$ is the i^{th} row of $\mathbf{X}$), and $\hat{\sigma}_p^2=\Big((q-p)\cdot(1,-\tilde{\mathbf{x}}^{(j)^T})\Omega(1,-\tilde{\mathbf{x}}^{(j)^T})^T\Big)^{-1}\sum_{t=1}^{q}\Big(Z_{tj}-\hat{\mathbf{b}}_t^T\tilde{\mathbf{x}}^{(j)}\Big)^2$ is an estimate of $\hat{\sigma}^2$ obtained from the prediction phase. An appropriate estimate of $\mathbf{x}^{(j)}$ is $\tilde{\mathbf{x}}^{(j)}$. An estimate of m_{22} is $\tilde{m}_{22}^{(j)}=\tilde{\mathbf{H}}^{(j)^T}(\mathbf{M}^{(j)}-\hat{\Sigma})\tilde{\mathbf{H}}^{(j)}$,

where $\tilde{\mathbf{H}}^{(j)}=(\mathbf{0},\mathbf{I}_p)^T-(1,-\tilde{\mathbf{x}}^{(j)^T})^T[(1,-\tilde{\mathbf{x}}^{(j)^T})\hat{\Sigma}(1,-\tilde{\mathbf{x}}^{(j)^T}]^{-1}(1,-\tilde{\mathbf{x}}^{(j)^T})\hat{\Sigma}_s$, $\hat{\Sigma}_s$ consists of columns 2 through p+1 of $\hat{\Sigma}$, and $\hat{\Sigma}=\hat{\sigma}^2\Omega$. See Thomas (1991) for discussion concerning where the normal approximation using $\hat{\Gamma}_j$ is accurate in this calibration context.

The joint distribution of the columns of $\tilde{\mathbf{X}}_0=\Big(\tilde{\mathbf{x}}_0^{(1)}, \tilde{\mathbf{x}}_0^{(2)},\ldots, \tilde{\mathbf{x}}_0^{(r)}\Big)$ is unknown. However, it is straightforward to derive the joint distribution of the $\tilde{\mathbf{X}}_0$ in the limit as $q\to\infty$ (or $\sigma^2\to 0$). The Appendix provides this limiting distribution.

4. JOINT MAXIMUM LIKELIHOOD ESTIMATION

When the estimated parameters of a multivariate calibration model are to be used repeatedly, responses from a batch of new samples (in $\mathbf{Y}_0$) can be used to estimate jointly the columns of $\mathbf{X}_0$ by maximum likelihood. This differs from the traditional approach where the estimation of a single column of $\mathbf{X}_0$ involves only the related column of $\mathbf{Y}_0$ and the estimated model parameters. As we shall see in Section 5, joint estimation could provide some advantages over individual estimation, especially under conditions where individual maximum likelihood estimation seems to perform poorly relative to least squares (see Thomas 1991).

Given the knowledge of the distribution of $\epsilon_{t\cdot}$, the joint MLE of $\mathbf{X}_0$ (and $\mathbf{B}$) can be

obtained by maximizing the log likelihood,

$$-(n/2)\cdot\log\left|2\pi\sigma^2\Omega_a\right|-(2\sigma^2)^{-1}\sum_{t=1}^{q}\left((Z_t,\hat{b}_t^T)-(z_t,b_t^T)\right)\Omega_a^{-1}\left((Z_t,\hat{b}_t^T)-(z_t,b_t^T)\right)^T,$$

with respect to X_0 and B. Amemiya and Fuller (1984) give the *joint* MLE of X_0, denoted here by $\tilde{\tilde{X}}_0$. Let $\rho_1 \geq \rho_2 \geq \ldots \geq \rho_{p+r}$ be the eigenvalues of $\Omega_a^{-\frac{1}{2}} M_a \Omega_a^{-\frac{1}{2}}$ and let $T = (T_1, T_2)$ be the matrix of corresponding orthonormal eigenvectors such that $\Omega_a^{-\frac{1}{2}} M_a \Omega_a^{-\frac{1}{2}} T_2 = T_2 R$,

where $M_a = q^{-1}\sum_{t=1}^{q} S_t S_t^T$, $S_t = (Z_{t1}, Z_{t2}, \ldots, Z_{tr}, \hat{b}_t^T)^T$, $R = \mathrm{diag}(\rho_{p+1}, \rho_{p+2}, \ldots, \rho_{p+r})$, and $\Omega_a^{-\frac{1}{2}}$ is the matrix square root of Ω_a^{-1}. Further, let $C = \Omega_a^{-\frac{1}{2}} T_2$. Then,

$$\tilde{\tilde{X}}_0 = \left(\tilde{\tilde{x}}^{(1)}, \tilde{\tilde{x}}^{(2)}, \ldots, \tilde{\tilde{x}}^{(r)}\right) = -C_{pr}C_{rr}^{-1}, \qquad [4.1]$$

where C_{rr} consists of the first r rows of C, while C_{pr} consists of the final p rows of C. Note that if $\bar{b} = \lim_{q\to\infty} q^{-1}\sum_{t=1}^{q} b_t$ and $\bar{m}_{22} = \lim_{q\to\infty} q^{-1}\sum_{t=1}^{q} b_t b_t^T$ is positive definite, $\tilde{\tilde{X}}_0$ exhibits strong consistency with respect to q (see Amemiya and Fuller 1984).

In general, the joint distribution of $\tilde{\tilde{X}}_0$ is unknown. With the assumptions given for consistency, Amemiya and Fuller (1984) and Fuller (1987), page 305, gave the distribution of $\tilde{\tilde{X}}_0$ in the limit as $q\to\infty$. Since by prior assumptions Σ_a is known to within a proportionality constant, $\Upsilon^{-\frac{1}{2}}\,\mathrm{vec}\,(\tilde{\tilde{X}}_0 - X_0) \xrightarrow{\ell}$ **Normal(0,I)**, as $q\to\infty$, where

$$\Upsilon = \begin{bmatrix} \Upsilon_{1,1} & \Upsilon_{1,2} & \cdots & \Upsilon_{1,r} \\ \Upsilon_{2,1} & \Upsilon_{2,2} & \cdots & \Upsilon_{2,r} \\ \vdots & \vdots & & \vdots \\ \Upsilon_{r,1} & \Upsilon_{r,2} & \cdots & \Upsilon_{r,r} \end{bmatrix} = \frac{\sigma^2}{q}\Psi\otimes\left(m_{22}^{-1} + \sigma^2 m_{22}^{-1}\Theta m_{22}^{-1}\right), \qquad [4.2]$$

$\Psi = \left(I_r, -X_0^T\right)\Omega_a\left(I_r, -X_0^T\right)^T$, $\Theta = \left\{\left(X_0, I_p\right)\Omega_a^{-1}\left(X_0, I_p\right)^T\right\}^{-1}$, and $\otimes$ is the Kronecker product.

In practice, neither σ^2, X_0, or m_{22}, which are elements of Υ, will be known. An estimator of Υ, $\hat{\Upsilon}$ can be obtained by replacing σ^2, X_0, or m_{22} in the expression for Υ with appropriate estimates. An estimate of σ^2 derived from both the calibration and prediction phases is $\hat{\sigma}^2 = \left((n+r)\cdot(q-p)-n\right)^{-1}\cdot\left\{\left(n\cdot(q-p-1)\right)\hat{\sigma}_c^2 + r\cdot(q-p)\,\hat{\sigma}_p^2\right\}$, where $\hat{\sigma}_c^2$ is from Section 3, and $\hat{\sigma}_p^2 = \left((q-p)\cdot r\right)^{-1}\sum_{t=1}^{q}\left(Z_t - \hat{b}_t^T\tilde{\tilde{X}}_0\right)\hat{\Psi}^{-1}\left(Z_t - \hat{b}_t^T\tilde{\tilde{X}}_0\right)^T$, where $\hat{\Psi} = \left(I_r, -\tilde{\tilde{X}}_0^T\right)\Omega_a\left(I_r, -\tilde{\tilde{X}}_0^T\right)^T$ (see Fuller 1987, page 294) . An appropriate estimate of X_0 is $\tilde{\tilde{X}}_0$. To estimate m_{22}, Fuller (1987, page 304) recommends using $\tilde{\tilde{m}}_{22} = \tilde{\tilde{H}}^T\left(M_a - \hat{\Sigma}_a\right)\tilde{\tilde{H}}$, where $\tilde{\tilde{H}} = \hat{\Sigma}_a^{-1}\left(\tilde{\tilde{X}}, I_p\right)^T\left[\left(\tilde{\tilde{X}}, I_p\right)\hat{\Sigma}_a^{-1}\left(\tilde{\tilde{X}}, I_p\right)^T\right]^{-1}$, and $\hat{\Sigma}_a = \hat{\sigma}^2\Omega_a$.

Approximate individual and joint confidence regions can easily be developed for various subsets of the $p \times r$ elements of $\mathbf{X_0}$. For instance, an approximate $(1-\alpha)$ confidence region for the first row of $\mathbf{X_0}$ is given by

$$\begin{bmatrix} 1 & 0 & . & . & 0 \end{bmatrix}\left(\tilde{\tilde{\mathbf{X}}}_0 - \mathbf{X_0}\right)\hat{\Upsilon}_s^{-1}\left(\tilde{\tilde{\mathbf{X}}}_0 - \mathbf{X_0}\right)^T \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leq \chi^2_{r;\,1-\alpha}, \text{ where}$$

$\hat{\Upsilon}_s$ is a subset of $\hat{\Upsilon}$ containing only those elements that are in rows $1,\ p+1,\ 2p+1,\ldots,$ $(r-1)p+1$, *and* columns $1,\ p+1,\ 2p+1,\ldots,\ (r-1)p+1$, of $\hat{\Upsilon}$. This particular subset of $\mathbf{X_0}$ relates to the first of the p underlying explanatory variables for all r samples.

Coincidentally (see Fuller 1987), the maximum likelihood estimate of $\left(z_t, \mathbf{b}_t^T\right)^T$, $t = 1, 2, \ldots, q$, modified by the information supplied by all of the new samples is

$$\left(\tilde{\tilde{z}}_t, \tilde{\tilde{\mathbf{b}}}_t^T\right)^T = \left(Z_t, \hat{\mathbf{b}}_t^T\right)^T\left[I - CC^T\Omega_a\right], \qquad [4.3]$$

The joint MLE of $\mathbf{b}_t$, $\tilde{\tilde{\mathbf{b}}}_t$, incorporates information from both the n calibration samples and r new samples. When $r = 1$ (i.e., individual maximum likelihood estimation) $\tilde{\tilde{\mathbf{b}}}_t = \tilde{\mathbf{b}}_t^{(1)}$ won't generally improve much on $\hat{\mathbf{b}}_t$. For large r, $\tilde{\tilde{\mathbf{b}}}_t$ can differ significantly from the various $\tilde{\mathbf{b}}_t^{(j)}$ and be a much better estimator of $\mathbf{b}_t$ (see Section 6). This is indicative of the potential improvement of joint maximum likelihood estimation over individual maximum likelihood estimation.

5. ASYMPTOTIC COMPARISON OF $\tilde{\tilde{\mathbf{X}}}_0$ AND $\tilde{\mathbf{X}}_0$

The purpose of this section is to identify situations where $\tilde{\tilde{\mathbf{X}}}_0$ and $\tilde{\mathbf{X}}_0$ can differ significantly. For this analysis, we will assume that the distributions of $\tilde{\tilde{\mathbf{X}}}_0$ and $\tilde{\mathbf{X}}_0$ are well approximated by the asymptotic results given earlier. In particular, the matrices, Γ_j and $\Upsilon_{j,j}$, which describe the asymptotic covariance structure among the elements of $\tilde{\mathbf{x}}^{(j)}$ and $\tilde{\tilde{\mathbf{x}}}^{(j)}$, respectively, will be used here.

The increase in asymptotic variance of the individual MLE when compared to that of the joint MLE can be represented by the non-negative matrix

$$\mathbf{Q_j} = \Gamma_j - \Upsilon_{j,j} = \frac{\sigma^4}{q}\mathbf{m}_{22}^{-1}\left(\left(\Omega_{22}\psi_j - \Lambda_j\Lambda_j^T\right) - \Theta\psi_j\right)\mathbf{m}_{22}^{-1}.$$

Under appropriate conditions, as $r \to \infty$, $\Theta \to \mathbf{0}$ so that $\mathbf{Q_j} \to \frac{\sigma^4}{q}\mathbf{m}_{22}^{-1}\left(\Omega_{22}\psi_j - \Lambda_j\Lambda_j^T\right)\mathbf{m}_{22}^{-1}$. For purposes of simplification, assume $\mathbf{x}^{(j)} = \mathbf{0}$, which implies that $\Lambda_j\Lambda_j^T = \mathbf{0}$. The difference, $\mathbf{Q_j}$, could be of practical significance if $\sigma^2\Omega_{22}\mathbf{m}_{22}^{-1}$ is large. This can be the case if: 1. the noise-to-signal ratio, $\sigma^2\mathbf{m}_{22}^{-1}$, is poor (large), 2. the number of calibration samples is relatively small, implying that Ω_{22} has relatively large elements, and 3. the specificity of the response variables is poor (i.e.,

$\mathbf{m}_{22}$ is ill-conditioned). If the number of new samples, r, is large and these three conditions are met, use of the joint MLE can offer a significant advantage.

To highlight the potential differences between joint and individual maximum likelihood estimation, we will consider the asymptotic performances ($q \to \infty$) of individual and joint maximum likelihood estimation in a spectroscopic context. Consider a hypothetical chemical system where two unconstrained components, when at unit concentrations, exhibit the absorbance spectra illustrated in Figure 1. The Gaussian spectral shapes portrayed in Figure 1 have means of 200 and 400 (arbitrary units), and a common standard deviation of 100. The response vector ($\mathbf{y}$) associated with samples containing both of these two components is assumed to follow Beer's law. That is, the t^{th} element of the q-dimensional spectrum is $y_t = a_t + (x_1, x_2)(b_{1t}, b_{2t})^T + e_t$, where x_1 and x_2 are the concentrations of the two chemical components and $(e_1, e_2, \ldots, e_q)^T$ are uncorrelated normally distributed measurement errors. Here the elements of the intercept vector, $(a_1, a_2, \ldots, a_q)$, are identically zero, but assumed to be unknown during calibration. The particular values of $\mathbf{b}_t = (b_{1t}, b_{2t})^T$ depend on q. For a given value of q, the set of frequencies used is $\{0, \Delta_q, 2\Delta_q, \ldots, (q-1)\Delta_q\}$, where $\Delta_q = \frac{600}{q-1}$. Figure 2 illustrates the values of $\{\mathbf{b}_t\}$ when q=20.

As q increases, the separation between adjoining frequencies decreases. As $q \to \infty$, $\{\mathbf{b}_t\}$ converge to the spectral shapes portrayed in Figure 1. It can be shown that $\lim_{q\to\infty} \mathbf{m}_{22} = \lim_{q\to\infty} q^{-1} \sum_{t=1}^{q} \mathbf{b}_t \mathbf{b}_t^T =$

$$\frac{1}{12\sqrt{\pi}} \begin{bmatrix} \Phi(4\sqrt{2}) - \Phi(-2\sqrt{2}) & e^{-1}\left\{\Phi(3\sqrt{2}) - \Phi(-3\sqrt{2})\right\} \\ e^{-1}\left\{\Phi(3\sqrt{2}) - \Phi(-3\sqrt{2})\right\} & \Phi(4\sqrt{2}) - \Phi(-2\sqrt{2}) \end{bmatrix} \simeq \begin{bmatrix} .0469 & .0173 \\ .0173 & .0469 \end{bmatrix}.$$

Conditions in which the spectra of the two chemical components are more overlapped were also considered (see Figure 3). In this case, the response variables have less specificity (the spectral shapes have means of 250 and 350, and a common standard deviation of 100), and $\lim_{q\to\infty} \mathbf{m}_{22} =$

$$\frac{1}{12\sqrt{\pi}} \begin{bmatrix} \Phi(3.5\sqrt{2}) - \Phi(-2.5\sqrt{2}) & e^{-.25}\left\{\Phi(3\sqrt{2}) - \Phi(-3\sqrt{2})\right\} \\ e^{-.25}\left\{\Phi(3\sqrt{2}) - \Phi(-3\sqrt{2})\right\} & \Phi(3.5\sqrt{2}) - \Phi(-2.5\sqrt{2}) \end{bmatrix} \simeq \begin{bmatrix} .0470 & .0366 \\ .0366 & .0470 \end{bmatrix}.$$

Two variations of $\mathbf{X}_f$ (columns 2 and 3 are assumed to be centered) were considered. In the first variation of $\mathbf{X}_f$ (denoted by $\mathbf{X}_{f_1}$), which relates to the case where relatively few calibration samples are used, $(\mathbf{X}_{f_1}^T \mathbf{X}_{f_1})^{-1} = \text{diag}(.2, 1, 1)$. In the second variation of $\mathbf{X}_f$ (denoted by $\mathbf{X}_{f_2}$),

which relates to the case where relatively many calibration samples are used, $(\mathbf{X}_{t_2}^T \mathbf{X}_{t_2})^{-1} = \text{diag}(.02, .1, .1)$. Two variations of σ^2 were also considered ($\sigma^2 \in \{.001, .01\}$). In total, the eight conditions defined by all combinations of the three two-level factors (spectral overlap, $\mathbf{X}_t$, and σ^2) were studied. The asymptotic performances of individual and joint maximum likelihood estimation are compared for five points in the plane, $\mathbf{x}^{(1)^T} \in \{(0, 0), (-.6, -.6), (-.6, .6), (.6, -.6), (.6, .6)\}$, which cover the assumed design space spanned by the centered values of $\mathbf{X}_t$.

The diagonal elements of Γ_1 (normalized by q) were computed for the eight conditions described above by using the limiting expressions for $\mathbf{m}_{22}$. Similarly, the diagonal elements of $\Upsilon_{1,1}$ were computed based on various assumptions. First to provide a benchmark, $\Upsilon_{1,1}$ was computed assuming that the model parameters, $\{a_t, b_t\}$, are known (i.e., $\Sigma_{22} = \mathbf{0}$). In this case, $\tilde{\tilde{\mathbf{x}}}^{(j)} = \tilde{\mathbf{x}}^{(j)}$, and $\Gamma_1 = \Upsilon_{1,1} = \Upsilon^0$, where $\Upsilon^0 = \text{block diag}\left\{\frac{\sigma^2}{q}\mathbf{m}_{22}^{-1}\right\}$. This approximates the performance of the maximum likelihood estimators when n (the number of calibration samples) is very large. The second variation of $\Upsilon_{1,1}$, denoted by Υ^1, relates to the case where the r new observations are numerous and sufficiently dispersed from the origin (i.e., $\Theta \approx \mathbf{0}$). The asymptotic covariance of $\tilde{\tilde{\mathbf{x}}}^{(1)}$ in this situation (assuming $\Theta = \mathbf{0}$) is $\Upsilon^1 = \frac{\sigma^2}{q}\psi_1 \mathbf{m}_{22}^{-1}$. This reflects the best possible performance of the joint maximum likelihood estimator in the general case when $\Sigma_{22} \neq \mathbf{0}$.

Other variations of $\Upsilon_{1,1}$ considered here assume that there are $r = 20$ new observations to predict. In many applications in analytical chemistry, the number of new observations to predict within a batch could easily exceed this number. Three cases are considered in order to provide some insight with regard to the effect of the range of the values in $\mathbf{X}_0$ on performance. In the first case, each element of columns 2-20 of $\mathbf{X}_0$ was obtained by independently sampling the uniform distribution on [-.9, .9], giving rise to Υ^{2a}. In the second case, columns 2-20 of $\mathbf{X}_0$ were obtained by sampling U[-.3, .3], giving rise to Υ^{2b}. In the third case, columns 2-20 of $\mathbf{X}_0$ were obtained by sampling U[-.9, .9] for the first row, and sampling U[-.3, .3] for the second row, giving rise to Υ^{2c}.

The asymptotic efficiencies of the individual and joint maximum likelihood estimators (with various assumptions on $\mathbf{X}_0$) relative to the performance achieved when all model parameters are known (i.e., $\Sigma_{22} = \mathbf{0}$) are given in Tables 1-5. For each estimator, the best performance is achieved when $\mathbf{x}^{(1)} = (0, 0)^T$ where, in fact, $\Upsilon^1 = \Upsilon^0$. Naturally, uncertainty in the model parameters causes poorer performance when $\mathbf{x}^{(1)}$ is distant from the center of the design space (e.g., $\mathbf{x}^{(1)} = (\pm .6, \pm .6)^T$). This can be an important issue when the uncertainty in the model parameters is relatively large (i.e., $\mathbf{X}_t = \mathbf{X}_{t_1}$). Also, apart from effect on the distance of $\mathbf{x}^{(1)}$ from the center of design space, the asymptotic performances of these estimators depend somewhat on

the orientation of $\mathbf{x}^{(1)}$ due to the overlap of the spectral features. Asymptotic performance is slightly better in some cases when $\mathbf{x}^{(1)} = (\alpha, -\alpha)^T$ versus when $\mathbf{x}^{(1)} = (\alpha, \alpha)^T$.

From Tables 1-5, note the hierarchy, $\Gamma_1(i,i) \geq \Upsilon^{2(abc)}(i,i) \geq \Upsilon^1(i,i)$, for each value of $\mathbf{x}^{(1)}$ for the eight conditions considered. This indicates the gains in asymptotic efficiency that are possible by using joint rather than individual maximum likelihood estimation when r is moderate and very large. When the response error variance, σ^2, is small (here $\sigma^2 = .001$), there is little difference between individual and joint maximum likelihood estimation. However, as presented in Tables 1-5, the joint maximum likelihood estimator can offer significant increases in efficiency when the response error variance is not trivial with respect to the underlying signal (here $\sigma^2 = .01$), especially when the response variables have poor specificity (here spectral features have severe overlap) and relatively few calibration samples are used (here $\mathbf{X}_f = \mathbf{X}_{f_1}$). This is precisely the condition where individual maximum likelihood estimation exhibits poor relative performance with respect to the classical least squares estimator (see Thomas 1991).

The asymptotic efficiency of $\tilde{\tilde{\mathbf{x}}}^{(j)}$ depends to some degree on the values in the other columns in $\mathbf{X}_0$. When the values are relatively highly dispersed from $(0, 0)^T$, the asymptotic efficiency of $\tilde{\tilde{\mathbf{x}}}^{(1)}$ is improved when compared to the case when values are relatively poorly dispersed (compare Tables 2 and 3). This difference is most notable when σ^2 is large, overlap is high, and the number of calibration samples is small. Note that the asymptotic performance of $\tilde{\tilde{\mathbf{x}}}^{(1)}$ given in Table 2 can be only marginally improved with additional new samples as A.R.E $(\Upsilon^{2a}(i,i))$ is not much smaller than A.R.E. $(\Upsilon^1(i,\ i))$. The asymptotic performance of $\tilde{\tilde{\mathbf{x}}}^{(1)}$, summarized in Table 4, is intermediate when compared to Table 2 (high dispersion) and Table 3 (low dispersion). Although, the second row of $\mathbf{X}_0$ is less dispersed than the first row of $\mathbf{X}_0$, A.R.E $(\Upsilon^{2c}(2,\ 2))$ is very close to the A.R.E. $(\Upsilon^{2c}(1,\ 1))$.

6. SIMULATION COMPARISON OF $\tilde{\tilde{\mathbf{X}}}_0$ AND $\tilde{\mathbf{X}}_0$

The previous analysis relates to the asymptotic case where $q \to \infty$. To get a feel for the difference between $\tilde{\tilde{\mathbf{X}}}_0$ and $\tilde{\mathbf{X}}_0$ with finite q (chosen here to be 100), a small simulation experiment was performed. We consider two conditions based on the setup described in the previous section where $\tilde{\tilde{\mathbf{X}}}_0$ and $\tilde{\mathbf{X}}_0$ are expected to perform differently ($\sigma^2 = .01$, $\mathbf{X}_f = \mathbf{X}_{f_1}$, with both low and high overlap).

For each of the two conditions studied, two values of $\mathbf{x}^{(1)}$ were considered; $\mathbf{x}^{(1)} = (0,\ 0)^T$ and $\mathbf{x}^{(1)} = (.6,\ .6)^T$. 1000 independent simulations were performed in each case. For each

simulation, the individual and joint MLE's involving $\mathbf{x}^{(1)}$ were obtained. In addition, $\mathbf{x}^{(1)}$ was estimated by least-squares (i.e. $\hat{\mathbf{x}}^{(1)} = \mathbf{M}_{22}^{-1}\,\mathbf{M}_{21}^{(1)}$). In the case of the joint MLE, a single realization of columns 2-20 of $\mathbf{X}_0$ was obtained by independently sampling the uniform distribution on [-.9, .9] 38 times. These simulations were carried out on an IBM-compatible personal computer using MATLAB© (Moler, et al., 1991).

Figures 4 and 5 summarize the distributions of the first elements of three estimators when $\mathbf{x}^{(1)} = (0,\ 0)^T$ and $\mathbf{x}^{(1)} = (.6,\ .6)^T$, respectively. Note that the distributions of the second elements of the three estimators are similar to the distributions of the first element because of symmetry and are not displayed. When the spectral overlap is relatively high, the joint MLE offers a significant reduction in variability when compared to the individual MLE; whereas when the spectral overlap is relatively low, the joint MLE offers a more modest reduction in variability. Significantly, wild values that occasionally appeared in the individual MLE when the overlap was high (maximum and/or minimum values of the first element of $\tilde{\mathbf{x}}^{(1)}$ were offscale in Figures 4 and 5) did not appear when using the joint MLE. For example, the maximum value of the first element of $\tilde{\mathbf{x}}^{(1)}$ exceeded 3.8, when $\mathbf{x}^{(1)} = (.6,\ .6)^T$ and the overlap was relatively high.

It is also interesting to compare the performance of the simpler least-squares estimator with the two MLE's for the cases considered. In these cases, the least-squares estimator, $\hat{\mathbf{x}}^{(1)}$, has smaller dispersion than the two MLE's. However, when $\mathbf{x}^{(1)} = (.6,\ .6)^T$ the least-squares estimator exhibits significant shrinkage towards $(0,\ 0)^T$. When $\mathbf{x}^{(1)} = (0,\ 0)^T$, the shrinkage towards $(0,\ 0)^T$ causes no problem.

These simulations indicate that the joint MLE is an improvement over the individual MLE by reducing dispersion while maintaining quasi-unbiasedness. When compared to the least-squares estimator, both MLE's offer relative unbiasedness while sacrificing dispersion. The bias associated with the least-squares estimator can be significant if the new $\mathbf{x}$ is far from the center of the $\mathbf{x}$'s in the training set.

It is also informative to compare various estimates of $\mathbf{B}$ that can arise during the calibration experiment (calibration and prediction). Here, for example, we will consider a single realization when $q = 100$, the overlap is relatively high (see Figure 3), $\sigma^2 = .01$, and $\mathbf{X}_f = \mathbf{X}_{f_1}$. Columns 2-20 were obtained as described earlier in this section while $\mathbf{x}^{(1)} = (0,\ 0)^T$. Figure 6 portrays the estimate of $\mathbf{B}$ based only on the calibration phase for one realization of the simulation (see equation 1.2). Figure 7 portrays the analogous estimate of $\mathbf{B}$ based only on the calibration phase and $\mathbf{y}^{(1)}$, corresponding to equation 3.2. (individual MLE). Figure 8 portrays the analogous

estimate of **B** based on the calibration phase and all columns of $\mathbf{Y}_0$, corresponding to equation 4.3 (joint MLE). The estimator of **B** obtained by joint maximum likelihood estimation incorporates all available information from both the n calibration samples and r new samples. As illustrated in Figures 5, 6, and 7, the estimator of **B** obtained by using all columns of $\mathbf{Y}_0$ can be much more precise than either of the other two estimators given. This advantage is present when $\frac{r+n}{n}$ is much larger than one.

7. CONCLUSION

When a classical multivariate calibration model is used repeatedly to predict some characteristic, maximum likelihood estimation involving the simultaneous use of responses from a batch of new samples can significantly improve prediction. This is especially true when the number of calibration samples is small, the response specificity is poor, and the new samples are numerous and well spread out. Furthermore, the limiting distribution of the joint maximum likelihood estimator, obtained directly from the errors-in-variables literature, can be used directly to develop approximate simultaneous confidence regions for $\mathbf{X}_0$.

The inherent batch nature of chemical assays makes joint estimation a potentially useful improvement in a number of calibration applications in analytical chemistry. The explicit nature of the causal model discussed in this paper makes it possible to take advantage of joint estimation. However, without an explicit model, as is the case of some of the popular methods in the chemometrics literature (e.g., partial least squares regression), it is probably more difficult to simultaneously incorporate the responses from a batch of new samples during prediction.

ACKNOWLEDGEMENTS

Useful comments by Yasuo Amemiya are gratefully acknowledged.

REFERENCES

Amemiya Y., and Fuller, W. A. (1984), "Estimation for the Multivariate Errors-in-Variables Model with Estimated Error Covariance Matrix," *The Annals of Statistics*, 12, 497-509.

Brown, P. J. (1982), "Multivariate Calibration," *Journal of the Royal Statistical Society*, Ser. B, 44, 287-308.

Brown, P. J., and Sundberg, R. (1987), "Confidence and Conflict in Multivariate Calibration," *Journal of the Royal Statistical Society*, Ser. B, 49, 46-57.

Carroll, R. J., Spiegelman, C. H., and Sacks, J. (1988), "A Quick and Easy Multiple-Use Calibration-Curve Procedure," *Technometrics*, 30, 137-141.

Denham, M. C., and Brown, P. J. (1991), "Regression, Sequenced Measurements, and Calibration," Technical report, University of Liverpool.

Fuller, W. A. (1987), *Measurement Error Models*, New York: John Wiley.

Martens, H., and Næs, T. (1989), *Multivariate Calibration*, Chichester: John Wiley.

Mee, R. W., Eberhardt, K. R., and Reeve, C. P. (1991), "Calibration and Simultaneous Tolerance Intervals for Regression," *Technometrics*, 33, 211-219.

Moler, C., Little, J., Bangert, S., and Kleiman, S. (1991), *386-MATLAB Users Guide*, Sherborn, Mass., The MathWorks Inc.

Næs, T. (1985), "Multivariate Calibration When the Error Covariance Matrix is Structured," *Technometrics*, 27, 301-311.

Oman, S. D. (1988), "Confidence Regions in Multivariate Calibration," *Annals of Statistics*, 16, 174-187.

Osborne, C. (1991), "Statistical Calibration: A Review," *International Statistical Review*, 59, 309-336.

Scheffé, H. (1973), "A Statistical Theory of Calibration," *Annals of Statistics*, 1, 1-37.

Thomas, E. V. (1991), "Errors-in-Variables Estimation in Multivariate Calibration," *Technometrics*, 33, 405-413.

APPENDIX - DERIVATION OF THE LIMITING DISTRIBUTION OF $\mathrm{vec}(\tilde{X}_0 - X_0)$

In derivation of the limiting distribution of $\tilde{x}_0$, Fuller (1987), page 129, provides an intermediate result that is used here. As a special case of this result, with Ω known to within a multiple,

$\tilde{x}_0^{(j)} - x_0^{(j)} = m_{22}^{-1} q^{-1} \sum_{t=1}^{q} g_{jt} + O_p(q^{-1})$, where

$g_{jt} = \left(\hat{b}_t v_{jt} - \frac{1}{\psi_j} \Lambda_j v_{jt}^2\right)$, and $v_{jt} = e_{0t}^{(j)} - \gamma_t - \delta_t^T x_0^{(j)}$.

To obtain the limiting distribution of $\mathrm{vec}(\tilde{X}_0 - X_0)$, we first find that

$\mathrm{Cov}(g_{jt}, g_{jt}) = \sigma^2 \psi_j b_t b_t^T + \sigma^4 \left\{\Omega_{22} \psi_j - \Lambda_j \Lambda_j^T\right\}$, and

$\underset{j \neq k}{\mathrm{Cov}}(g_{jt}, g_{kt}) = \sigma^2 (\psi_{j,k} - 1) b_t b_t^T + \sigma^4 \left\{\Omega_{22}(\psi_{j,k} - 1) + \Lambda_j \Lambda_k^T + 2\Lambda_j \Lambda_k^T \left(\frac{\psi_j - 1}{\psi_j}\right)\left(\frac{\psi_k - 1}{\psi_k}\right)\right.$

$\left. - 2\Lambda_j \Lambda_k^T \left\{\left(\frac{\psi_j - 1}{\psi_j}\right) + \left(\frac{\psi_k - 1}{\psi_k}\right)\right\}\right\}$.

Let $g_t = \mathrm{vec}(g_{1t}, g_{2t}, \ldots, g_{mt})$. From previous assumptions, we know that the g_t are *i.i.d.* random variables with $E[g_t] = 0$. Therefore, using the central limit theorem we find that as $q \to \infty$, $\Gamma^{-\frac{1}{2}} \mathrm{vec}(\tilde{X}_0 - X_0) \xrightarrow{\mathcal{L}}$ **Normal**$(0, I)$, where

$$\Gamma = \begin{bmatrix} \Gamma_{1,1} & \Gamma_{1,2} & \cdots & \Gamma_{1,m} \\ \Gamma_{2,1} & \Gamma_{2,2} & \cdots & \Gamma_{2,m} \\ \vdots & \vdots & & \vdots \\ \Gamma_{m,1} & \Gamma_{m,2} & \cdots & \Gamma_{m,m} \end{bmatrix},$$

$$\Gamma_{j,j} = \frac{\sigma^2}{q} \psi_j m_{22}^{-1} + \frac{\sigma^4}{q} m_{22}^{-1} (\Omega_{22} \psi_j - \Lambda_j \Lambda_j^T) m_{22}^{-1},$$

$$\Gamma_{j,k \neq j} = \frac{\sigma^2}{q} (\psi_{j,k} - 1) m_{22}^{-1} + \frac{\sigma^4}{q} m_{22}^{-1} \left\{\Omega_{22}(\psi_{j,k} - 1) + \Lambda_j \Lambda_k^T + 2\Lambda_j \Lambda_k^T \left(\frac{\psi_j - 1}{\psi_j}\right)\left(\frac{\psi_k - 1}{\psi_k}\right)\right.$$

$$\left. - 2\Lambda_j \Lambda_k^T \left\{\left(\frac{\psi_j - 1}{\psi_j}\right) + \left(\frac{\psi_k - 1}{\psi_k}\right)\right\}\right\} m_{22}^{-1}, \text{ where } \psi_{j,k} = \left(1, -x_0^{(j)T}\right) \Omega \left(1, -x_0^{(k)T}\right)^T,$$

$$\psi_j = \left(1, -x_0^{(j)T}\right) \Omega \left(1, -x_0^{(j)T}\right)^T, \text{ and } \Lambda_j = \left(\Omega_{21} - \Omega_{22} x_0^{(j)}\right).$$

Table 1- Comparison of the diagonal elements of Υ^0 and Υ^1 for various values of (x_1, x_2).

$$\text{A.R.E.}(\Upsilon^1(i, i)) = \Upsilon^0(i, i)/\Upsilon^1(i, i);\ i = 1,2$$

σ^2	$X_{\dagger}$	Overlap	$q\Upsilon^0(i, i)$	A.R.E. $(\Upsilon^1(i, i))$ $x^{(1)} = (0, 0)^T$	A.R.E. $(\Upsilon^1(i, i))$ $x^{(1)} = (\pm .6, \pm .6)^T$
.01	$X_{\dagger_1}$	High	.649	1	.62
.001	$X_{\dagger_1}$	High	.0649	1	.62
.01	$X_{\dagger_2}$	High	.552	1	.93
.001	$X_{\dagger_2}$	High	.0552	1	.93
.01	$X_{\dagger_1}$	Low	.296	1	.62
.001	$X_{\dagger_1}$	Low	.0296	1	.62
.01	$X_{\dagger_2}$	Low	.252	1	.94
.001	$X_{\dagger_2}$	Low	.0252	1	.94

Table 2- A.R.E. $(\Upsilon^{2a}(i, i)) = \Upsilon^0(i, i)/\Upsilon^{2a}(i, i)$ for various values of $x^{(1)}$.

$$\text{A.R.E.}(\Upsilon^{2a}(1, 1)) = \text{A.R.E.}(\Upsilon^{2a}(2, 2))$$

σ^2	$X_{\dagger}$	Overlap	$x^{(1)}$		
			$(0, 0)^T$	$(-.6, -.6)^T$ or $(.6, .6)^T$	$(-.6, .6)^T$ or $(.6, -.6)^T$
.01	$X_{\dagger_1}$	High	.92	.57	.58
.001	$X_{\dagger_1}$	High	.99	.62	.62
.01	$X_{\dagger_2}$	High	.96	.89	.89
.001	$X_{\dagger_2}$	High	1.00	.93	.93
.01	$X_{\dagger_1}$	Low	.97	.61	.61
.001	$X_{\dagger_1}$	Low	1.00	.62	.62
.01	$X_{\dagger_2}$	Low	.98	.92	.92
.001	$X_{\dagger_2}$	Low	1.00	.93	.93

Table 3- A.R.E. $(\Upsilon^{2b}(i, i)) = \Upsilon^{0}(i, i)/\Upsilon^{2b}(i, i)$ for various values of $\mathbf{x}^{(1)}$.

A.R.E. $(\Upsilon^{2b}(1, 1)) =$ A.R.E. $(\Upsilon^{2b}(2, 2))$

σ^2	$\mathbf{X}_{\dagger}$	Overlap	$\mathbf{x}^{(1)}$		
			$(0, 0)^T$	$(-.6, -.6)^T$ or $(.6, .6)^T$	$(-.6, .6)^T$ or $(.6, -.6)^T$
.01	$\mathbf{X}_{\dagger_1}$	High	.65	.41	.45
.001	$\mathbf{X}_{\dagger_1}$	High	.95	.60	.60
.01	$\mathbf{X}_{\dagger_2}$	High	.92	.86	.87
.001	$\mathbf{X}_{\dagger_2}$	High	.99	.93	.93
.01	$\mathbf{X}_{\dagger_1}$	Low	.85	.54	.55
.001	$\mathbf{X}_{\dagger_1}$	Low	.99	.62	.62
.01	$\mathbf{X}_{\dagger_2}$	Low	.98	.91	.91
.001	$\mathbf{X}_{\dagger_2}$	Low	1.00	.93	.93

Table 4- A.R.E. $(\Upsilon^{2c}(i, i)) = \Upsilon^{0}(i, i)/\Upsilon^{2c}(i, i)$ for various values of $\mathbf{x}^{(1)}$.

A.R.E. $(\Upsilon^{2c}(1, 1))$ [A.R.E. $(\Upsilon^{2c}(2, 2))$]

σ^2	$\mathbf{X}_{\dagger}$	Overlap	$\mathbf{x}^{(1)}$		
			$(0, 0)^T$	$(-.6, -.6)^T$ or $(.6, .6)^T$	$(-.6, .6)^T$ or $(.6, -.6)^T$
.01	$\mathbf{X}_{\dagger_1}$	High	.75 [.71]	.47 [.45]	.50 [.48]
.001	$\mathbf{X}_{\dagger_1}$	High	.97 [.97]	.60 [.59]	.61 [.60]
.01	$\mathbf{X}_{\dagger_2}$	High	.93 [.93]	.87 [.87]	.87 [.87]
.001	$\mathbf{X}_{\dagger_2}$	High	1.00 [1.00]	.93 [.93]	.93 [.93]
.01	$\mathbf{X}_{\dagger_1}$	Low	.92 [.86]	.58 [.55]	.59 [.56]
.001	$\mathbf{X}_{\dagger_1}$	Low	.99 [.98]	.62 [.62]	.62 [.62]
.01	$\mathbf{X}_{\dagger_2}$	Low	.98 [.98]	.92 [.91]	.92 [.91]
.001	$\mathbf{X}_{\dagger_2}$	Low	1.00 [1.00]	.93 [.93]	.93 [.93]

Table 5- A.R.E. $(\Gamma_1(i, i)) = \Upsilon^0(i, i)/\Gamma_1(i, i)$ for various values of $\mathbf{x}^{(1)}$.

A.R.E. $(\Gamma_1(1, 1))$ = A.R.E. $(\Gamma_1(2, 2))$

σ^2	X_{t}	Overlap	$\mathbf{x}^{(1)}$		
			$(0, 0)^T$	$(-.6, -.6)^T$ or $(.6, .6)^T$	$(-.6, .6)^T$ or $(.6, -.6)^T$
.01	X_{t_1}	High	.53	.34	.40
.001	X_{t_1}	High	.92	.57	.59
.01	X_{t_2}	High	.92	.86	.86
.001	X_{t_2}	High	.99	.93	.93
.01	X_{t_1}	Low	.78	.49	.52
.001	X_{t_1}	Low	.97	.61	.61
.01	X_{t_2}	Low	.97	.91	.91
.001	X_{t_2}	Low	1.00	.93	.93

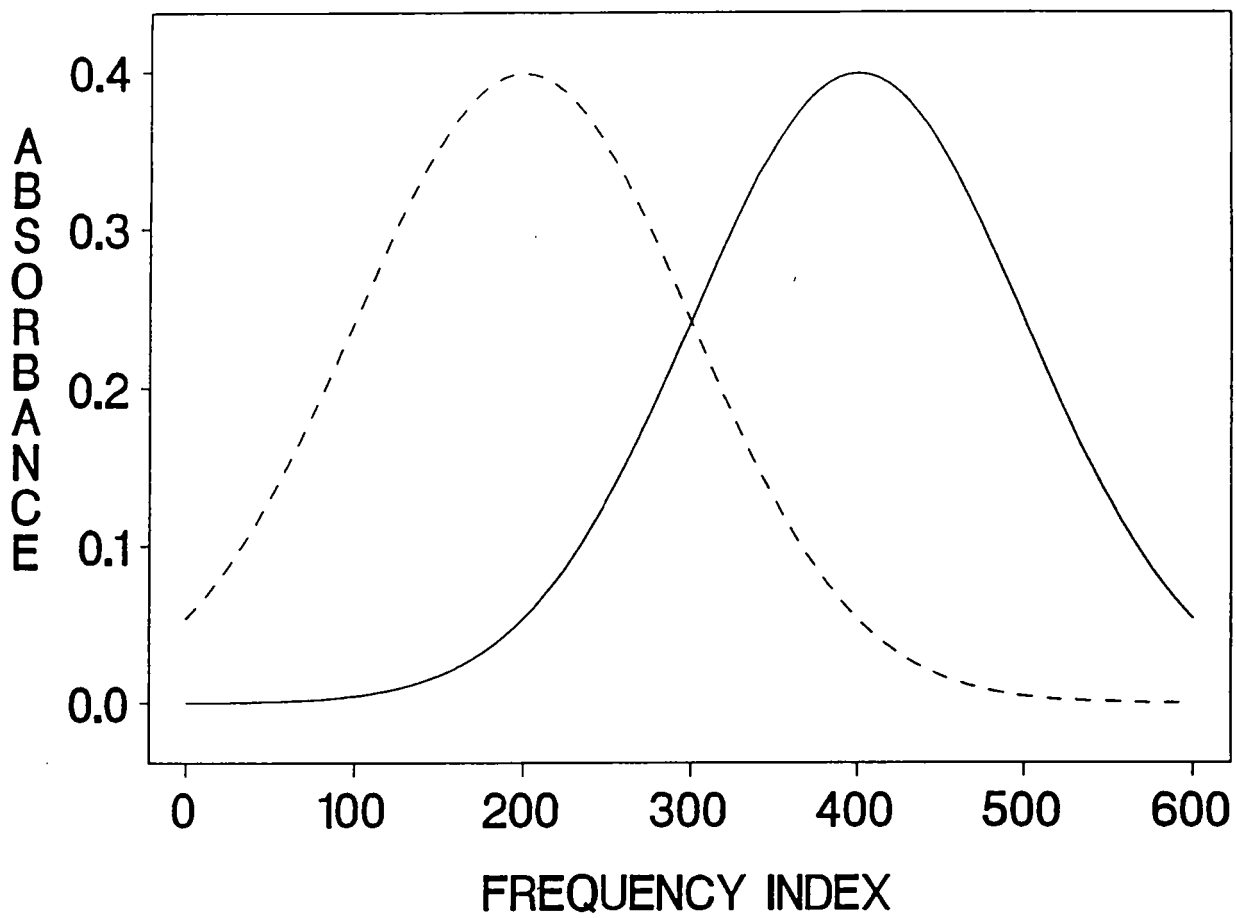


Figure 1- Hypothetical Spectra When Components Are At Unit Concentration. The spectra of components 1 and 2 are represented by a dashed curve and a solid curve, respectively. The frequency index is given in arbitrary units from 0 to 600.

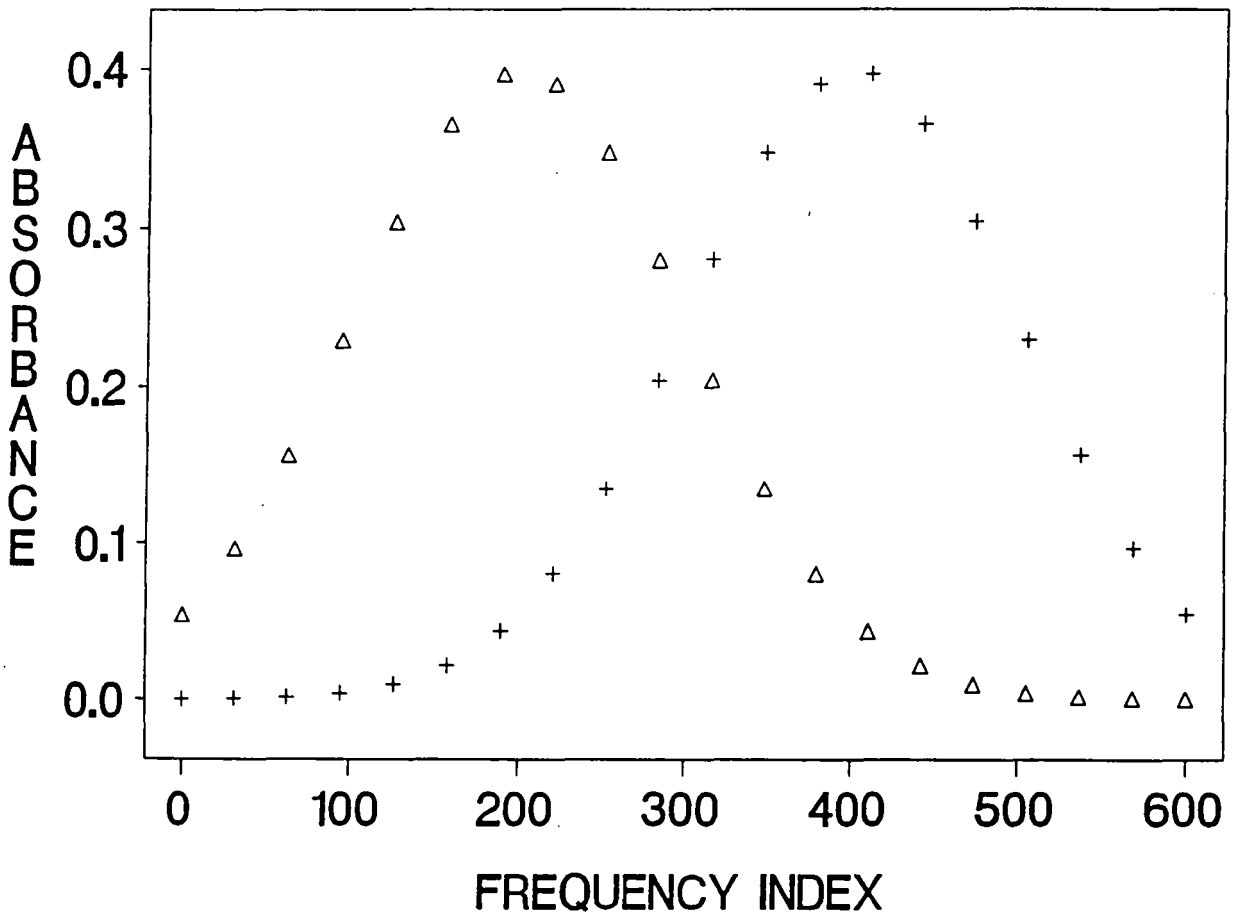


Figure 2- The Values of B When q=20. The elements of the first row of B are given by the ordinates of the Δ's. The elements of the second row of B are the ordinates of the + 's.

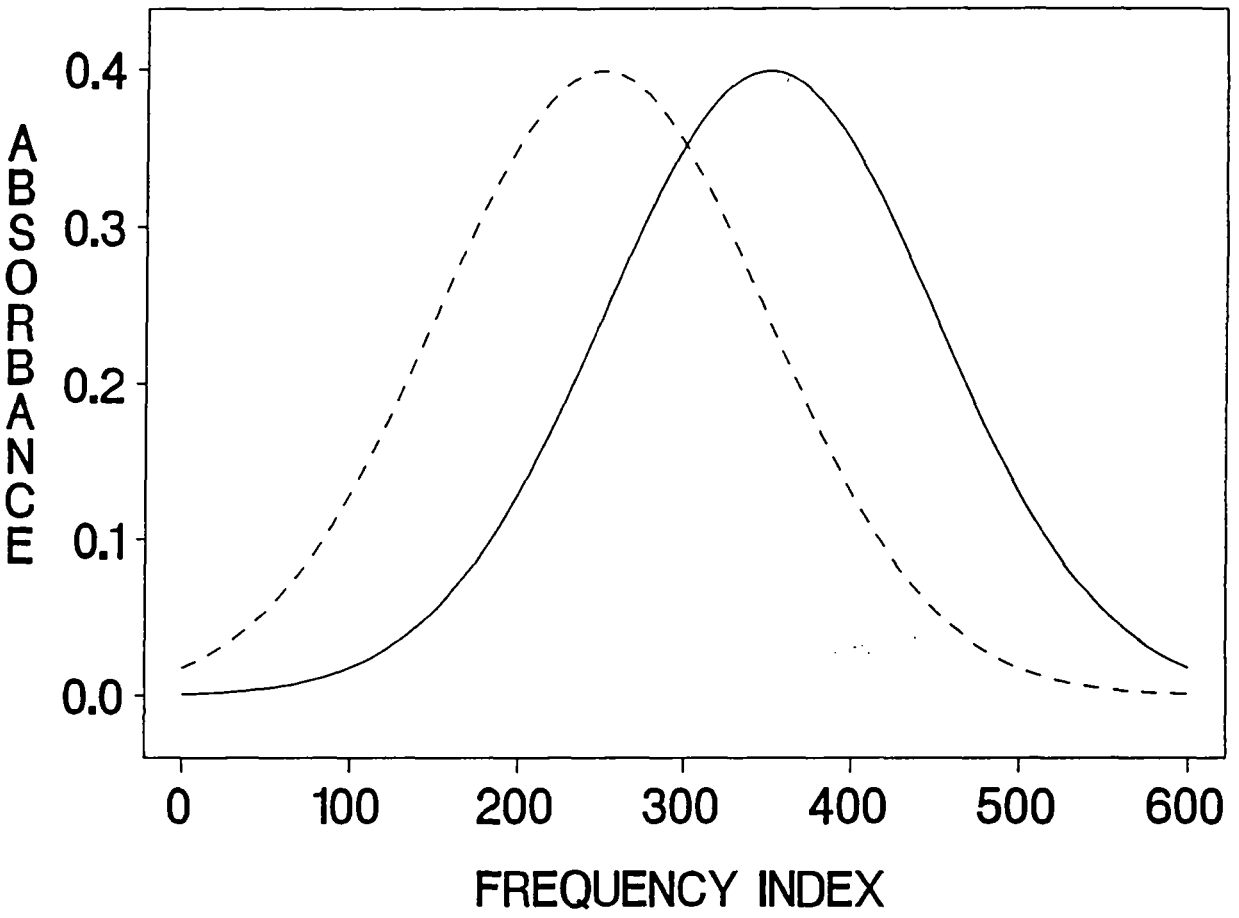


Figure 3- Hypothetical Spectra With Severe Overlap.

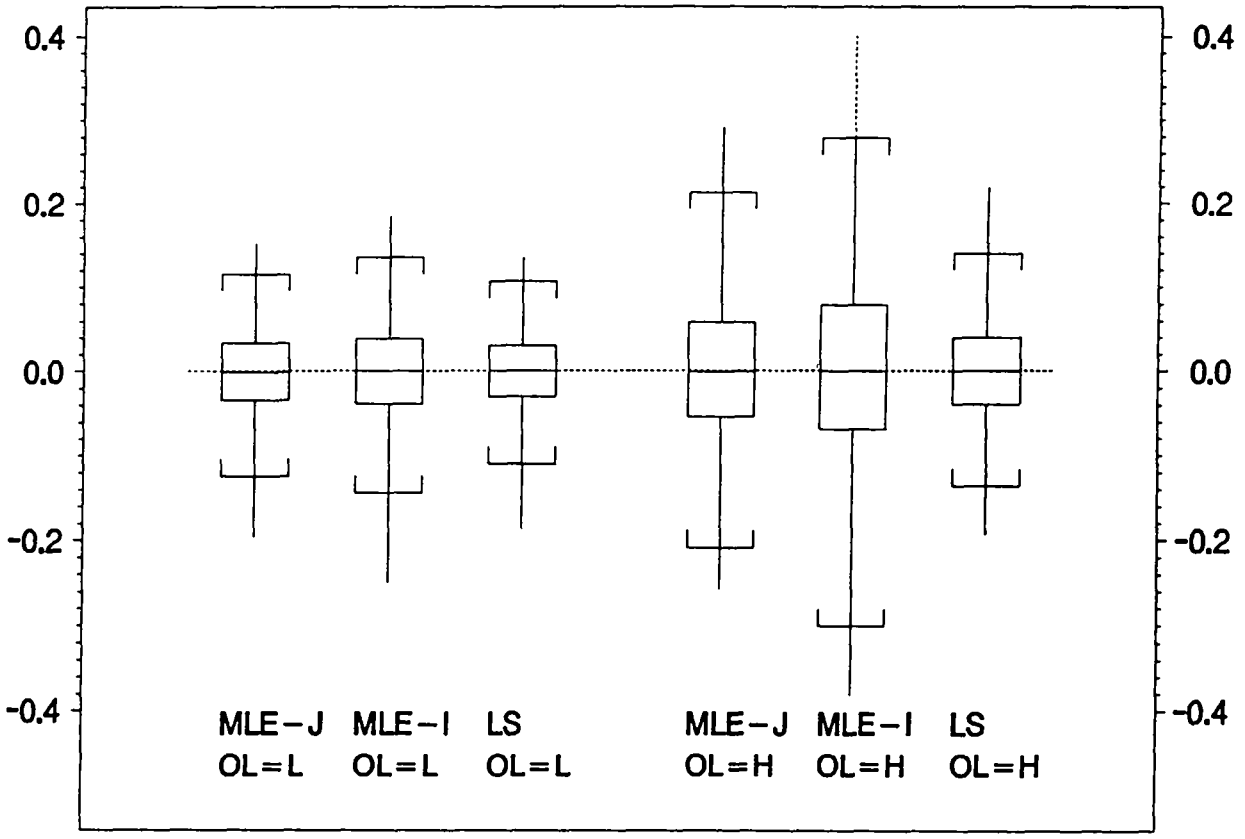


Figure 4- Box Plots of the distributions of the three estimators of the first element of $\mathbf{x}^{(1)} = (0, 0)^T$. The order statistics indicated are the minimum, 10^{th}, 250^{th}, median, 751^{st}, 991^{st}, and maximum. The simulation condition is indicated by the spectral overlap (OL). OL=L refers to the conditions of Figure 1 (relatively low overlap), while OL=H refers to the conditions of Figure 3 (relatively high overlap).

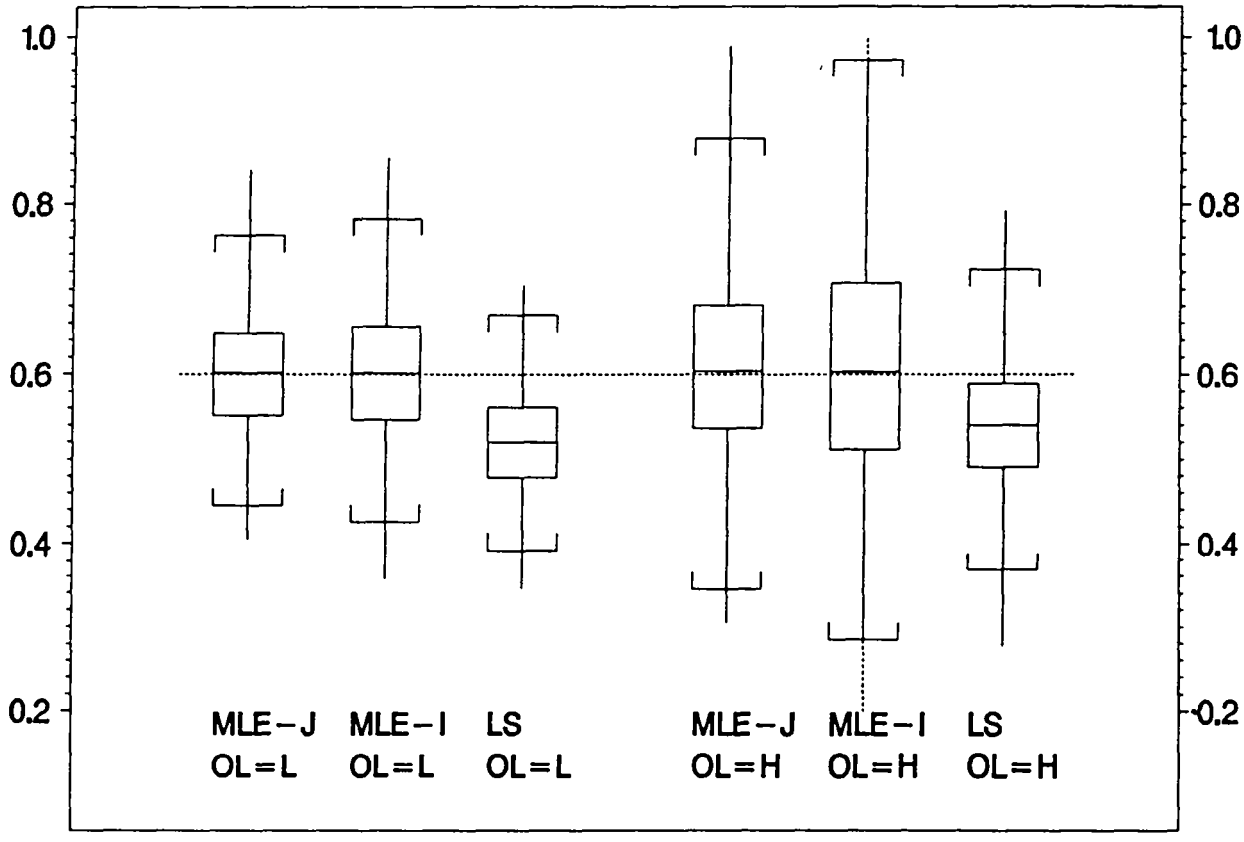


Figure 5- Box Plots of the distributions of the three estimators of the first element of $\mathbf{x}^{(1)} = (.6, .6)^T$.

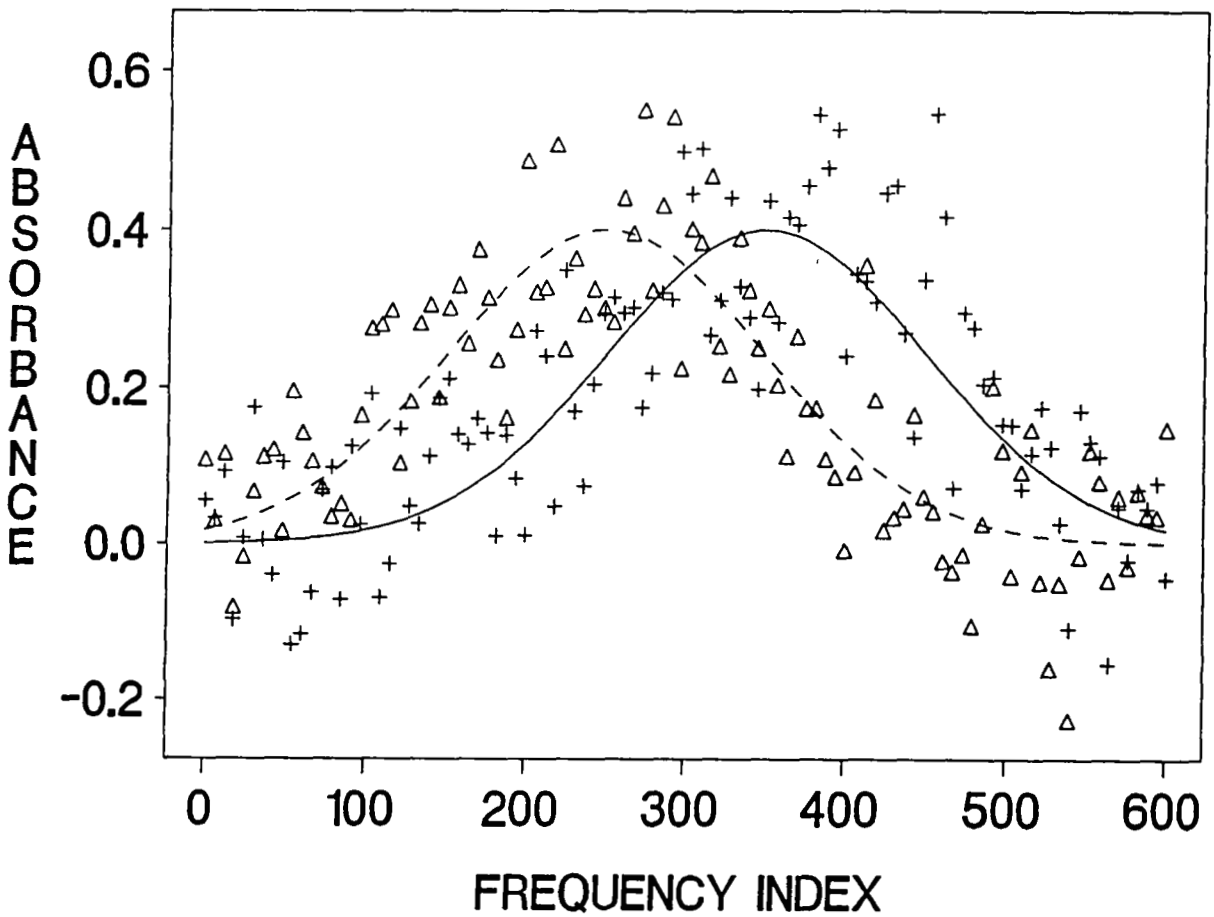


Figure 6- Estimates of the first (△) and second (+) rows of B based only on the calibration phase. The spectra that defines the first (----) and second (—) rows of B are overlaid for comparison.

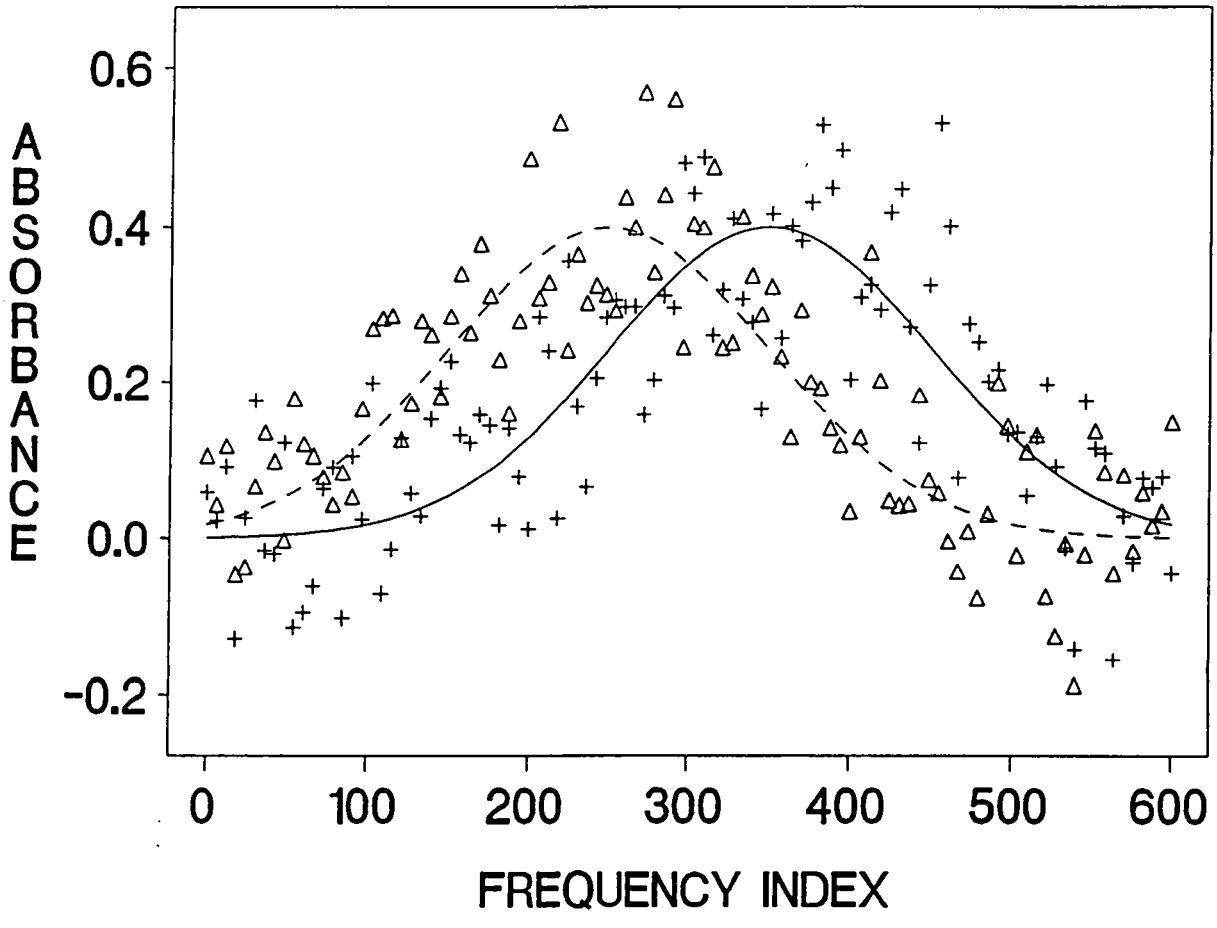


Figure 7- Estimates of the first (△) and second (+) rows of B based on individual maximum likelihood estimation (see equation 3.2).

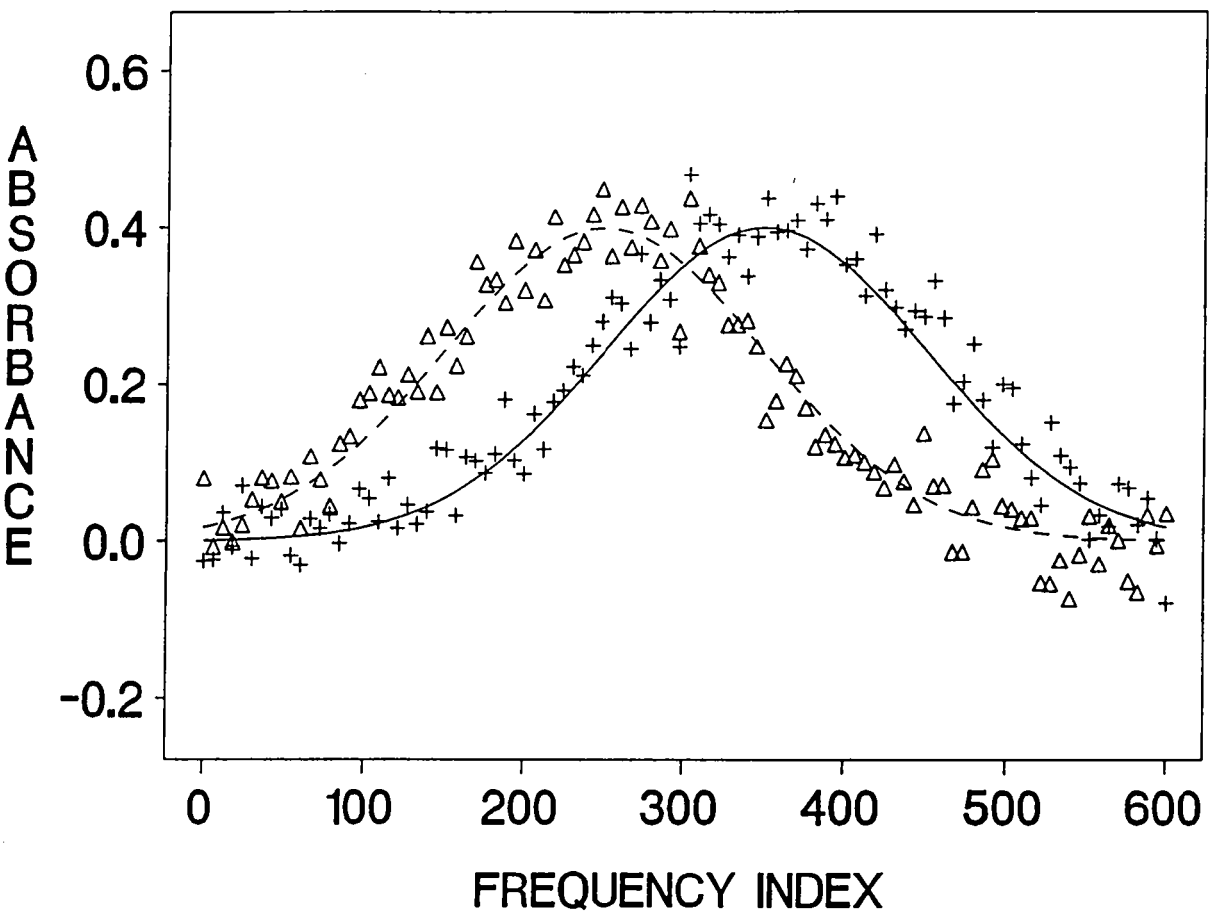


Figure 8- Estimates of the first (Δ) and second (+) rows of B based on joint maximum likelihood estimation (see equation 4.3).

DISTRIBUTION:

2 Rio Grande Medical Technologies
Attn: M. R. Robinson
R. Rowe
915 Camino de Salud, Box 603
Albuquerque, New Mexico 87131-5271

1 Iowa State University
Department of Statistics, Snedecor Hall
Attn: Professor Y. Amemiya
Ames, Iowa 50011

1	MS 0343	M. K. Alam, 1823
1	MS 0343	J. A. Borders, 1823
1	MS 0343	D. M. Haaland, 1823
1	MS 0829	K. V. Diegert, 12323
1	MS 0829	R. G. Easterling, 12323
30	MS 0829	E. V. Thomas, 12323 (30)
1	MS 9018	Central Technical Files, 8523-2
5	MS 0899	Technical Library, 13414
1	MS 0619	Technical Publications, 13416
10	MS 0100	Document Processing, 7613-2